

研究简报

# 汉语语音识别系统评估<sup>1)</sup>

王仁华 倪晋富

(中国科学技术大学 合肥 230027)

**关键词:** 语音识别,性能评价,语音数据库.

## 1 引言

汉语语音识别系统评估,是指运用科学的方法和技术手段,来评定不同的识别系统和算法之间的优劣. 这项研究对改进和完善现有系统设计,提高系统性能,引导汉语语音识别技术向期望的方向发展,以及加快产业化进程都有重要的意义. 全面的评价汉语语音识别系统的整体性能,是很困难的课题,目前无论在理论上,还是在实践方法上,都尚缺乏公认的标准和手段. 1992、1993年6月国家科委组织了两次汉语语音识别系统评测,取得了一定的效果,为进一步开展这项工作奠定了良好的基础,本文对此作一简单介绍.

## 2 测试的基本内容和方法

两次语音识别系统测试,是按照预先确定的《语音识别系统测试大纲》进行,测试内容包括:特定人的全音节测试、特定人的文本输入测试和特定人及非特定人的词组测试. 全音节识别是汉语语音识别的一个重要特点,单音节输入测试的目的,是测定系统在没有加入语法约束情况下,单音节识别能力,而文本输入测试是对系统中语音识别和字处理能力的一种综合测试. 单音节

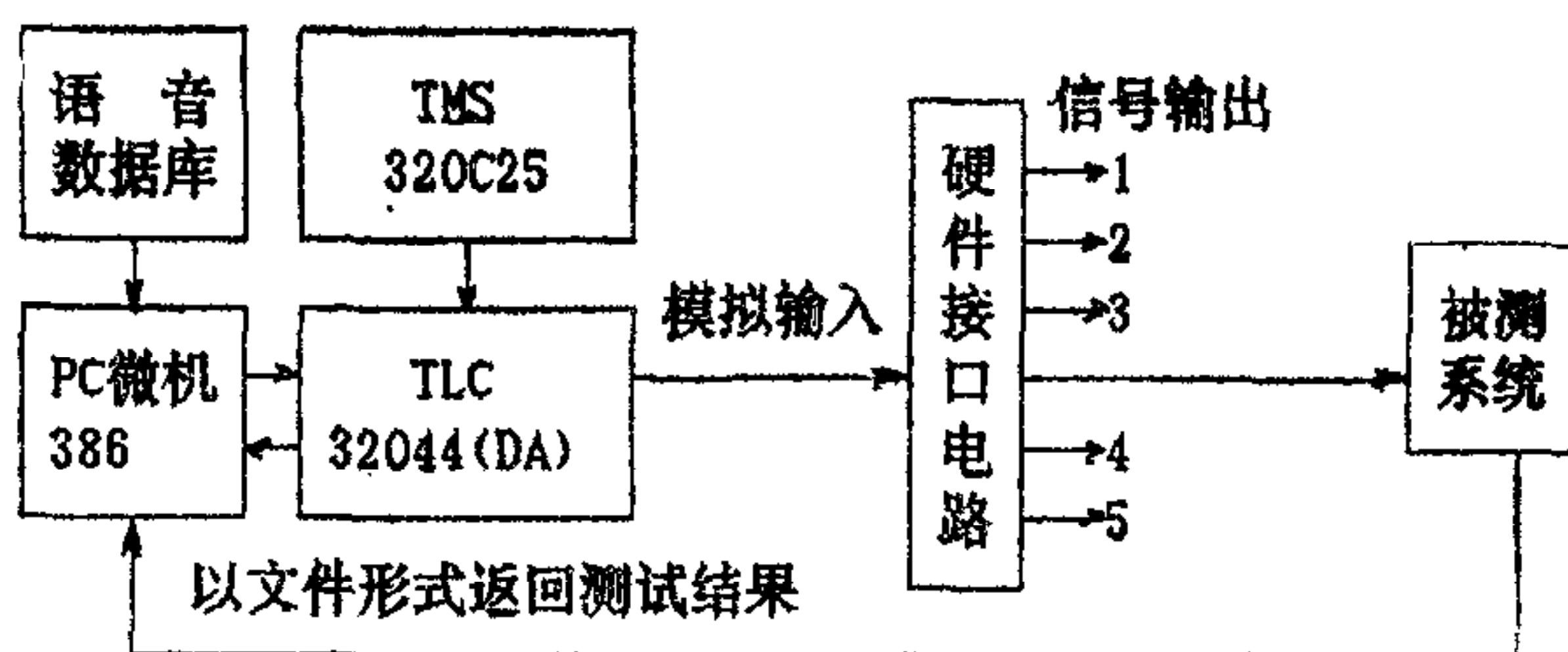


图1 基于数据库的测试系统框图

测试的主要指标包括,训练指标(训练遍数,每遍训练字数,训练时间等),识别率指标(主选音节正识率,主选无调音、1至4个候选音正识率、四声判决正确率),单音节测试语料覆盖了汉语全部无调单音节,文本输入测试语料是20条选自政治、经济、科技、文化等领域中的句子,测试指标是以语音方式正确输入这些句子所需要的时间. 词组测试是由参加评测系统的单位提供测试词表,测试指标类同单音节.

1) 本文得到863项目及自然科学基金项目资助.  
本文于1992年8月25日收到

测试过程中采用了两种测试方式.方式 1,在安静的办公室环境下,由说标准普通话的播音员,通过麦克风进行现场训练和测试;方式 2,基于语音数据库的半自动现场测试,图 1 给出该方式测试系统的基本框图.

这两种方式各有其特点,在方式 1 中,发音人会对系统进行一定配合,如适当调整音量、音长和嘴离话筒的距离和方位等,比较接近实验室的测试环境和实验性能,它的主要缺点是有一些影响识别性能的因素不易控制,难以保证在完全相同的条件下进行各系统间的性能比较.方式 2,是用预先记录的语音样本进行训练和测试,训练和测试模块按相应的控制文件,从数据库中搜索语音数据样本,接口系统将  $D/A$  输出的语音信号衰减到与麦克风输出相当的幅度,均衡的输出给被测系统,完成训练和测试.这种方法降低了人对测试过程的干预程度,可以在一致的环境下,进行多个系统测试,不仅有节约时间、便于各系统间比较和测试可重复等优点,同时,由于测试语音是预先录制的,发音人对识别系统并无先验知识,发音具有一定的自由度,测试中还可以使用不同发音时间的语音,以及模拟不同的噪声环境,因此,测试结果较方式 1 更能反映系统的实际性能,测试也更具科学性,实际测试结果也证实了上述分析.

### 3 总结

通过对两次测试总结,提出以下几点看法供大家参考:

#### 1. 全音节识别系统的问题和前景

以孤立音节识别为基础,辅以词法等知识的汉字语音输入系统仍然是研究主流,特定人汉语全音节识别系统已做到实时识别,但还存在一些问题值得深入考虑:

- 单音识别率仍需要进一步提高,其中,关键是要提高声母识别率.统计表明,全音节主选正识率在 80% 左右,进一步对主选错误分析统计表明,声母正确而韵母错误的仅占 15.4%,而韵母正确声母错误要占 61.9%.

- 系统对发音要求太严,说话人要合机器“胃口”,系统抗噪声能力较差.

- 孤立字识别方式大大限制了语音识别系统的应用,其应用前景不容乐观.

#### 2. 词组识别进步大,前景看好.

特定人和非特定人的词组识别有了长足进步,表现在主选正识率提高,词汇量增大,特定人大词汇量(万条以上)词组识别已达 90% 的主选正识率.由于发音人可以以较大自由度的连续发音,使得词组识别方式输入文本似乎比单音输入有更光明的前途.

#### 3. 增强系统的抗噪声能力和顽健性,是语音识别走出实验室的关键.建立有效的语言模型,增强系统对语音信号可变性和声学环境可变性的自适应能力(包括发音人因素和环境噪声的影响),是把语音识别推向实用化迫切需要克服的难关.

汉语语音识别系统评估的研究在我国刚刚起步,建立科学的评测系统,需要加强基础研究.不仅需要研究科学的评价标准和评价方法,以衡量、评定不同识别系统及不同处理方法之间的优劣,预测其在不同使用条件下的性能;而且,由于语音信号强烈地依赖于发音人的个人特征和周围环境,研制和应用标准的语音数据库,是保证评价系统具有科学性的基本条件之一.

## ASSESSMENT OF CHINESE SPEECH INPUT SYSTEMS

WANG RENHUA    NI JINFU

*(University of Science and Technology of China, Hefei, China 230027)*

**Key words:** Speech recognition, assessment evaluation, Speech database.